DOCUMENT RESUME

ED 093 905 TM 003 720

AUTHOR Penfield, Douglas A.; Koffler, Stephen L.
TITLE A Comparison of Some K-Sample Nonparametric Tests for Scale.
PUB DATE [Apr 74]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974); Appendixes may reproduce poorly

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Analysis of Variance; Comparative Analysis; Computer Programs; *Nonparametric Statistics; Sampling

ABSTRACT

Four nonparametric alternatives to the parametric Bartlett test are presented for handling the K-sample equality of variance problem. The two sample Siegel-Tukey Test, Mood Coefficient of Alienation Test, and Klotz Test are extended to the multisample situation by the methods of Puri. A fourth alternative involving a Q-statistic procedure developed by Chatterjee is also presented. These K-sample scale tests are illustrated and compared with respect to computational methodology and asymptotic relative efficiency for specified alternatives. (Author)

ED 093905

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 003 220

13.18

BEST COPY AVAILABLE

A COMPARISON OF SOME K-SAMPLE NONPARAMETRIC TESTS FOR SCALE

Douglas A. Penfield and Stephen L. Koffler
Rutgers University

Paper presented at the annual meeting of the American Educational Research Association on April 17, 1974

Introduction

The term "scale" in statistics is synonymous with dispersion or variability of data points. When a behavioral scientist finds himself in need of determining whether differences in the scale (variance) of K (K > 2) samples are due to differences in the scale of the parent populations or due to chance variations which were to be expected among random samples from the same population, it is customary for him to apply tests such as those developed by Bartlett (1947) and Cochran (1941) to the set of data. Both tests assume an underlying distribution of scores which is approximately normally distributed and tend to be quite sensitive to departures from normality. Thus, when the normality assumption is suspect, the researcher is forced to seek out nonparametric analogues to test his hypothesis of interest.

For the case when K = 2, the nonparametric tests for scale most frequently mentioned in the literature are the Siegel-Tukey test, the Mood test and the Normal Scores test. A description and comparison of these tests relative to the behavioral sciences was prepared by Penfield (1972).

Puri (1964) has suggested a generalized test statistic for the nonparametric homogeneity of variance situation. Using the Puri statistic, the two sample tests developed by Siegel and Tukey (1960), Mood (1954) and Klotz (1962) can be extended to cover K - samples. This paper contrasts the methods of development of these three tests for scale and applies then to an example. In addition, a generalized Q statistic developed by Chatterjee (1965) and providing another distribution-free alternative to the K - sample scale problem is discussed.

Generalized Test Development

Let $X_{i,j}$ ($i = 1, \ldots, n_j$; $j = 1, \ldots, K$) represent K independent samples drawn from populations having continuous cumulative distribution function, $F_j(X) = F_j(\theta_j(X - u))$, where θ_j and u are real numbers. θ_j is the scale parameter for F_j. u is the common median over all F_j and can be considered to be zero. The hypothesis to be tested is $H_o: \theta_1 = \theta_2 = \ldots = \theta_k$ against the alternative H_1: H_o if false. Combine the $n_1 + n_2 + \ldots + n_k = N$ scores of the K samples. Using an indicator variable, $Z^{(j)}_{N,i}$, let $Z^{(j)}_{N,i} = 1$ if the i^{th} smallest observation from the combined sample is from the j^{th} sample and 0 otherwise. The generalized Puri statistic is defined as

$$L = \sum_{j=1}^{k} n_j \frac{(T_{N,j} - \mu_{N,j})^2}{A^2}$$

where $\mu_{N,j}$ and A^2 are normalizing constants and $n_j T_{N,j} = \sum_{i=1}^{n_j} E_{N,i} Z^{(j)}_{N,i}$

The statistic L has a limiting distribution which is non-central chi-square with K - 1 degrees of freedom when $\mu_{N,j} = E(T_{N,j})$ and $A^2 = \text{Var}(T_{N,j})$. Hence, since $n_j T_{N,j} = \sum_{i=1}^{n_j} E_{N,i} Z^{(j)}_{N,i}$, the following equalities hold

$$(1) \quad E(T_{N,j}) = \bar{E}_{N,i}$$

$$(2) \quad \text{Var}(T_{N,j}) = \frac{(N-n_j)}{n_j N(N-1)} \sum_{i=1}^{N} (E_{N,i} - \bar{E}_{N,i})^2$$

Mood Test for K-Samples

This test requires replacing the scores in the pooled sample of size N with their corresponding ranks. Next determine the sum of squared rank deviations about the mean, (N+1)/2. Using a derivative of this procedure, $E_{N,i} = \left(\frac{i}{N+1} - .5\right)^2$. Making appropriate substitutions, $E(T_{N,j}) = \frac{N-1}{12(N+1)}$ and $\text{Var}(T_{N,j}) = \frac{(N-n_j)(N^2-4)}{180\, n_j (N+1)^3}$. This leads to a generalized Puri statistic synonomous with Mood's procedure

and is written as

$$(3)\quad L = 180 \sum_{j=1}^{k} \frac{((N+1)^3(\sum_{i=1}^{N}(\frac{i}{N+1} - .5)^2 Z_{N,i}^{(j)} - \frac{n_j(N-1))^2}{12(N+1)})}{(N-n_j)(N^2-4)}$$

Siegel-Tukey Test for K-Samples

In this instance the pooled data is replaced by a reordering of the ranks (i) from 1 to N. The lowest score gets a rank of 1 while the largest and second largest scores receive ranks of 2 and 3 respectively. Returning to the lower end of the distribution, the second smallest and third smallest scores are assigned ranks of 4 and 5. This process is repeated until all scores are accounted for. The Puri statistic generalizes to the Siegel-Tukey model when $E_{N,i} = i$, $E(T_{N,j}) = \frac{N+1}{2}$ and $\text{Var}(T_{N,j}) = \frac{(N-n_j)(N+1)}{12\, n_j}$. For this situation, the generalized Puri statistic reduces to

$$(4)\quad L = 12 \sum_{j=1}^{k} \frac{(\sum_{i=1}^{N} iZ_{N,i} - n_j(N+1))^2}{(N-n_j)(N+1)}$$

Normal Scores Test for K-Samples

This test represents an attempt to reconstruct the F test using expected normal order statistics, $E(V^i)$, in place of the original scores, Let $V^1 < V^2 < \ldots < V^N$ be an ordered sample of size N from a normal distribution. If V^i is the i^{th} ranked score in the combined sample of size N (since values are conditional upon N), then $E(V^i)$ is the expected value of the score in the i^{th} position assuming the score has come from a standard normal population. The term $E(V^i)$ acts as a distance measure in much the same way that Z scores express relative distance in a normal distribution.

To extend the Normal Scores test to K-samples using the generalized Puri model, let $E_{N,i} = (E(V^i))^2$, $E(T_{N,j}) = \sum_{i=1}^{N} \frac{(E(V^i))^2}{N}$

and $\text{Var}(T_{N,j}) = \frac{(N-n_j)}{n_j N(N-1)} \left(\sum_{i=1}^{N} (E(V^i))^4 - \frac{1}{N} \left(\sum_{i=1}^{N} (E(V^i))^2 \right)^2 \right)$.

Substituting these values into the generalized Puri statistic yields

$$(5)\quad L = N(N-1) \sum_{j=1}^{K} \frac{\left(\left(\sum_{i=1}^{N} (E(V^i))^2 Z_{N,i} - \frac{n_j}{N} \sum_{i=1}^{N} (E(V^i))^2 \right)^2 \right)}{(N-n_j)\left(\sum_{i=1}^{N} (E(V^i))^4 - \frac{1}{N} \left(\sum_{i=1}^{N} (E(V^i))^2 \right)^2 \right)}$$

Chatterjee's Q test

A fourth alternative for the homogeneity of scale problem which is not amenable to the Puri statistic was developed by Chatterjee in 1966. This test, Q, is distribution-free and is based on U-statistics. As was true with the generalized Puri statistic, Q is asymptotically distributed as chi-square with K-1 degrees of freedom.

To compute the statistic, rank the combined data and replace each score, X_{ij}, by its corresponding rank, R_{ij}. The Q statistic is defined as

$$(6)\quad Q = \frac{45}{N(N+1)} \sum_{j=1}^{K} \frac{n_j(n_j-1)(N-n_j)^2(T_j - 1/3)^2}{(n_j + 3)}$$

where $T_j = \frac{2 \sum_{i=1}^{n_j} (i - \frac{n_j+1}{2}) R_{ij} - \frac{1}{6} n_j(n_j^2-1)}{(2^{n_j})(N-n_j)}$

R_{ij} is the rank of the i^{th} score in the j^{th} sample.

It is readily apparent that the computations involved in ascertaining the Chatterjee Q statistic are quite formidable for even a small number of samples having only a few observations. Therefore, to assist the experimenter, a Fortran IV computer program was written

to compute the Q statistic and its related chi-square significance level (see Appendix I).

The following hypothetical example is cited to illustrate the various K-sample nonparametric tests for the homogeneity of scale problem.

Hypothetical Example

Forty second grade girls known to have a low expectation of success on intellectual tasks and a high anxiety about performance in school were randomly assigned to four experimental conditions of stress. The dependent variable of interest was the amount of time (in minutes) they would continue to attempt to solve a difficult problem.

The experimenter felt that score variability would vary across groups depending upon stress level. The four groups were ordered such that the first group received the least stress while the fourth group operated under the greatest amount of stress.

The hypothesis under test, H_o, was one of no difference in scale between the four groups. Let alpha = 0.05. Data necessary for computing the generalized Puri statistic relative to the Mood, Siegel-Tukey, and Normal Scores procedures as well as Chatterjee's Q and Bartlett's B are presented in Table 1. Since there are four groups, H_o will be rejected in all cases if the test statistic is greater than $X^2_3(.95) = 7.815$.

TABLE I

Time Data for the Puri analogues to the Mood, Siegel-Tukey and Normal Scores tests.

Group 1

($n_1 = d$)

Time (min)	Rank (i)	Siegel-Tukey Rank	$E(V^i)$
20	10	20	-.7099
21	13	25	-.4859
22	15	29	-.3498
23	16	32	-.2842
23	17	33	-.2199
23	18	36	-.1564
25	22	38	.0936
26	23	35	.1564
28	26	30	.3498
211	160	278	-1.6063

Group 2

($n_2 = 13$)

Time (min)	Rank (i)	Siegel-Tukey Rank	$E(V^i)$
17	5	9	-1.2033
18	6	12	-1.0833
19	7	13	-.9772
19	9	17	-.7926
22	14	28	-.4169
25	21	39	.0312
26	24	34	.2199
28	25	31	.2842
30	28	26	.4859
30	29	23	.5574
32	31	19	.7099
33	33	15	.8811
35	36	10	1.2033
334	268	276	-.1004

Group 3

(n_3 = 11)

Time (min)	Rank (i)	Siegel-Tukey Rank	$E(V^i)$
11	1	1	-2.1608
15	3	5	-1.5170
19	8	16	-.8811
21	11	21	-.6318
21	12	24	-.5574
24	19	37	-.0936
28	27	27	.4169
32	32	18	.7926
34	35	11	1.0833
39	38	6	1.5170
41	39	3	1.7531
285	225	169	-.2788

Group 4

(n_4 = 7)

Time (min)	Rank (i)	Siegel-Tukey Rank	$E(V^i)$
12	2	4	-1.7531
17	4	8	-1.3437
25	20	40	-0.0312
31	30	22	.6318
34	34	14	.9772
37	37	7	1.3437
42	40	2	2.1608
198	167	97	1.9855

The data necessary for computing the L, Q and B statistics will now be presented.

Mood

$$E_{N,i} = \left(\frac{i}{N+1} - .5 \right)^2$$

Group	$\sum_{i=1}^{N} E_{N,i} Z_{N,i}^{(j)}$
1	0.163
2	0.877
3	1.214
4	0.916

$$L = \frac{180\,(41)^3}{1596}\,(0.00976 + 0.00087 + 0.00403 + 0.00395)$$

$$= 144.7$$

Siegel-Tukey

$$E_{N,i} = i$$

Group	$\sum_{i=1}^{N} E_{N,i} Z_{N,i}^{(j)}$
1	278
2	276
3	169
4	37

$$L = 12\,(6.878 + 0.082 + 2.685 + 1.598)$$

$$= 134.9$$

Normal Scores

$$E_{N,i} = (E(V^i))^2$$

Group	$\sum_{i=1}^{N} (E(V^i))^2 Z_{N,i}$
1	1.1716
2	7.7835
3	15.8155
4	12.7085

$$L = 40(39)(0.0305 + 0.0128 + 0.0187 + 0.0205)$$
$$= 128.7$$

Chatterjee's Q

Using the computer program outlined in Appendix I, Q was found to equal 39.7.

The printout of pertinent information used in computing Q is presented in Appendix II.

Bartlett's B

$$B = \frac{2.30259}{C}\left((N-K)\log_{10}MS_e - \sum_{j=1}^{K}(n_j-1)\log_{10}S_j^2\right)$$

where $$C = 1 + \frac{1}{3(K-1)}\left(\sum_{j=1}^{K}\frac{1}{n_j-1} - \frac{1}{N-K}\right)$$

$$= 1 + \frac{1}{9}(.475 - .028) = 1.05$$

Group	n_j	S_j^2	$\log_{10}S_j^2$	$(n_j-1)\log_{10}S_j^2$
1	9	6.278	0.7978	6.3824
2	13	38.397	1.5842	19.0104
3	11	94.691	1.9763	19.7630
4	7	117.905	2.0716	12.4296
				57.5854

$$B = \frac{2.30259(36(1.7793) - 57.5854) + 14.2}{1.05}$$

In all instances, the hypothesis, H_o, of no scale differences among groups is rejected. There is a great similarity among the magnitude of the values computed under Puri's generalized L statistic. Essentially, it yields approximately equivalent results under the three different approaches.

Chatterjee's Q statistic appears more conservative than the Puri test and the Bartlett results show it to be the most conservative of all the tests under consideration for this example. It should be noted, however, that when the normality condition is satisfied, the parametric Bartlett test is the most powerful test. The data presented in the example is not normal and this could partially account for the vast difference in results between the parametric test and the nonparametric tests.

Asymptotic Relative Efficiency

Puri (1964) has noted that the efficiency of the L statistic agrees with the results obtained by Klotz (1962) for the two-sample situation. Klotz has compared the relative efficiency of the M, S-T, and NS tests for a specified number of distributions. For scores drawn from distributions with sharp tails (exponential, rectangular, etc.), the NS test is preferred to S-T and is equally as effective as M. When the distribution of scores has heavy tails (Cauchy, etc.), the S-T test is more efficient for testing equality of scale. For normally distributed data, Bartlett's test has the greatest relative efficiency. Assuming normality, the asymptotic relative efficiency of M to B is 0.76, of S-T to B is 0.61 and of NS to B is 1.0.

As for Chatterjee's Q, its efficiency relative to B is identical to that of Mood's M. It is most effective in detecting scale differences between groups when the data is other than normal. When the data is approximately normal, the Normal Scores test proposed by Klotz is more efficient that Q. There is no single nonparametric test for scale which is best for all alternatives.

REFERENCES

Bartlett, M. S. (1937). "Properties of Sufficiency and Statistical Tests," Proceedings of the Royal Society, 160, 268-282.

Chatterjee, S. K. (1966). "A Multi-Sample Nonparametric Scale Test Based On U Statistic," Calcutta Statistical Association Bulletin, 58, 109-119.

Cochran, W. G. (1941). "The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total," Annals of Eugenics, 11, 47-52.

Hajek, Jaroslav, (1969). A Course In Nonparametric Statistics, San Francisco: Holden Day.

Klotz, J. H. (1962). "Nonparametric Tests for Scale," Annals of Mathematical Statistics, 33, 495-512.

Mood, A. M. (1954). "On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests," Annals of Mathematical Statistics, 25, 514-522.

Puri, M. L. (1964). "Asymptotic Efficiency of a Class of C-Sample Test," Annals of Mathematical Statistics, 35, 102-121.

Puri, M. L. (1971). Nonparametric Methods in Multivariate Analysis, New York: John Wiley and Sons, Inc.

Puri, M. L. (1965). "On Some Tests of Homogeneity of Variances," Annals of the Institute of Statistical Mathematics, 17, 323-330.

Penfield, D. A. (1972). "A Comparison of Some Nonparametric Tests for Scale," Paper presented at the annual meeting of the American Educational Research Association.

Siegel, S. and Tukey, J. W. (1960). "A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples," Journal of the American Statistical Association, 55, 429-445.

```
FORTRAN IV G LEVEL 19                    MAIN                    DATE = 71...
0001                  DIMENSION N(10),X(10,100),R(200),RK(200),T(10)
0002                  DATA N1/0/,NN/0/,K/0/,Q/0.0/
                C
                C     READ THE NUMBER OF SAMPLES AND THE SAMPLE SIZE OF EACH
                C
0003                  READ(5,1)NS,(N(I),I=1,NS)
0004                  DO 2 I=1,NS
0005                  NT=NT+N(I)
0006                2 NN=MAXO(NN,N(I))
                C
                C     READ EACH ELEMENT OF EACH SAMPLE
                C
0007                  DO 3 I=1,NS
0008                3 READ(5,4)(X(I,J),J=1,NS)
0009                  XNT=NT
0010                  DO 5 J=1,NS
0011                  XN=N(J)
0012                  XN=NN
                C
                C     DETERMINE THE RANK OF EACH ELEMENT OF THE COMBINED SAM
                C
0013                  DO 6 I=1,NN
0014                  K=K+1
0015                  R(K)=X(I,J)
0016                6 CONTINUE
0017                5 CONTINUE
0018                  CALL RANKR(R,RK,NT)
0019                  K=0
0020                  DO 11 I=1,NS
0021                  XN=N(I)
0022                  NN=XN
                C
                C     DETERMINE THE COEFFICIENTS FOR THE T(I) STATISTIC
                C
0023                  COEFF=2./XN/(XN-1.)
0024                  COEFF=COEFF/(XNT-XN)
0025                  SUBT=XN*(XN**2-1.)/6.
0026                  DO 8 J=1,NN
                C
                C     DETERMINE THE VALUE OF THE T(I) STATISTICS
                C
0027                  K=K+1
0028                      T(I)=(J-(XN+1.)/2.)*RK(K)
0029                8 CONTINUE
0030                      T(I)=COEFF*(2.*T(I)-SUBT)
0031               11 Q=Q+XN*(XN-1.)*(XNT-XN)**2/(XN+3.)*(T(I)-.333)**2
                C
                C     DETERMINE THE VALUE OF THE Q STATISTIC
                C
0032                  Q=Q*6./XNT/(XNT-1.)
0033                  XNDF=NS-1.
0034                  NDF=XNDF
                C
                C      DETERMINE THE SIGNIFICANCE LEVEL OF THE Q STATISTIC.
                C     A CHI-SQUARE STATISTIC WITH DEGREES OF FREEDOM EQUAL TO
                C     OF SAMPLES MINUS 1.
```

```
C
      SIG=PROB(1.0,1000.0,Q/XNDF)
         KA=0
         K=0
      WRITE(6,654)
         DO 900 I=1,NS
      K=KA+1
         NK=N(I)
         KA=KA+NK
           WRITE(6,14)I,(N(L),L=K,KA)
      WRITE(6,15)(RK(L),L=K,KA)
  900 CONTINUE
      WRITE(6,456)
           DO 901 I=1,NS
  901      WRITE(6,22)I,T(I)
      WRITE(6,100)Q,NDF,SIG
    1 FORMAT(11(I2,1X))
    4 FORMAT('0F5.0)
   14 FORMAT(///5X,'SAMPLE',I2,1X,'DATA',10X,20F5.0)
   15 FORMAT(///16X,'RANK',10X,20F5.0)
   22 FORMAT(///3X,'SAMPLE',1X,I1,1X,'T(',I1,') = ',F8.6)
  100 FORMAT(////' THE CHATTERJEE STATISTIC IS ',F10.5,///' WITH '
     *I2,' DEGREES OF FREEDOM.'///' THIS VALUE IS SIGNIFICANT AT
     *F7.4,' LEVEL.')
  456 FORMAT(/////7X,'VALUES FOR T(I)')
  654 FORMAT(////49X,'RAW DATA AND RANKED DATA MATRIX')
      STOP
      END
```

```
C
C     FUNCTION ROUTINE FOR DETERMINING THE SIGNIFICANCE LEVEL
C     Q STATISTIC
C
      FUNCTION PROB (DA, DB, FR)
      PROB = 1.0
      IF(DA * DB * FR .EQ. 0.0) RETURN
      IF(FR. LT. 1.0) GO TO 5
      A = DA
      B = DB
      F = FR
      GO TO 10
    5 A = DB
      B = DA
      F = 1.0 / FR
   10 AA = 2.0 / (9.0 * A)
      BB = 2.0 /(9.0 * B)
      Z = ABS(((1.0 - BB) * F**0.333333 - 1.0 + AA) / SQRT(BB * F
     1 **0.666667 + AA))
      IF (B . LT. 4.0) Z = Z * (1.0 + 0.08 * Z**4 / B**3)
      PROB = 0.5 / (1.0 + Z * (0.196854 + Z * (0.115194 + Z *
     1 (0.000344 + Z * 0.019527 )))) **4
      IF(FR. LT. 1.0) PROB = 1.0 - PROB
      RETURN
      END
```

```
      C
      C     SUBROUTINE FOR RANKING THE COMBINED SAMPLE
      C
0001        SUBROUTINE RANKR(A,R,N)
0002        DIMENSION A(200),R(200)
0003        DO 10 I=1,N
0004     10 R(I)=0.0
0005        DO 100 I=1,N
0006        IF(R(I))20,20,100
0007     20 SMALL=0.0
0008        EQUAL=0.0
0009        X=A(I)
0010        DO 50 J=1,N
0011        IF(A(J)-X)30,40,50
0012     30    SMALL=SMALL+1.0
0013           GOTO50
0014     40    EQUAL=EQUAL+1.0
0015           R(J)=-1.0
0016     50 CONTINUE
0017        IF(EQUAL-1.0)60,60,70
0018     60 R(I)=SMALL+1.0
0019        GOTO100
0020     70 P=SMALL+(EQUAL+1.0)*0.5
0021        DO 90 J=1,N
0022        IF(R(J)+1.0)90,80,90
0023     80    R(J)=P
0024     90 CONTINUE
0025    100 CONTINUE
0026        RETURN
0027        END
```

APPENDIX II

SAMPLE 1 DATA	20.	21.	22.	23.	23.	23.	25.	26.	28.			
RANK	10.	12.	15.	17.	17.	17.	21.	24.	26.			
SAMPLE 2 DATA	17.	18.	19.	19.	22.	25.	26.	28.	30.	30.	32.	33.	35.
RANK	5.	6.	8.	8.	15.	21.	24.	26.	29.	29.	32.	33.	36.
SAMPLE 3 DATA	11.	15.	19.	21.	21.	24.	28.	32.	34.	39.	41.		
RANK	1.	3.	8.	12.	12.	19.	26.	32.	35.	38.	39.		
SAMPLE 4 DATA	12.	17.	25.	31.	34.	37.	42.						
RANK	2.	5.	21.	30.	35.	37.	40.						

VALUES FOR T(I)

SAMPLE 1 T(1) = 0.0769

SAMPLE 2 T(2) = 0.0323

SAMPLE 3 T(3) = 0.1066

SAMPLE 4 T(4) = 0.2655

THE CHATTERJEE STATISTIC IS 39.65608

WITH 3 DEGREES OF FREEDOM.

THIS VALUE IS SIGNIFICANT AT THE 0.0006 LEVEL.